DOCUMENT RESUME

ED 157 709 SE 024 513

AUTHOR Searls, Donald T.; And Others
TITLE Change in National Science Achievement: Some Analysis Problems.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
PUB DATE 75
NOTE 19p.; Contains occasional light and broken type

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; *Achievement; *Educational Research; *Reports; *Science Education; *Statistical Analysis; Surveys; Testing
IDENTIFIERS *National Assessment Educational Progress

ABSTRACT

This document explores some of the problems associated with understanding the data of a report completed by the National Assessment of Educational Progress (NAEP). The report deals with science achievement among young Americans over a period of four years. The procedure used by NAEP in assessing educational achievement and the factors involved in the selection of data is explained. A general problem related to change measures used by NAEP is explicated and approaches that might clarify possible factors associated with observed results are explored. Two kinds of data adjustment for science achievement measures are illustrated. One is aimed at removing masquerading and double counting due to disproportionate distributions of sub-populations. The other is aimed at estimating the impact of shifts of these distributions over time and the shifts in masquerading that might change real performance. (GA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Theodore Pratt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED157709

"Change In National Science Achievement: Some Analysis Problems"

Donald T. Searls, Robert C. Larson, Susan W. Sherman

National Assessment of Educational Progress
The Education Commission of the States

SE 024 513

"Change In National Science Achievement: Some Analysis Problems"

Donald T. Searls, Robert C. Larson, Susan W. Sherman

National Assessment of Educational Progress
The Education Commission of the States

INTRODUCTION

The National Assessment of Educational Progress (NAEP) has recently completed a report* that documents observed changes in Science achievement over a four-year period. Science will be assessed again in 1977 and subsequent years so that eventually information will be available on the long term trends in Science achievement in this country.

This article will explore some of the problems in understanding the underlying structure of the observed data and the complexities of pursuing analysis for change data.

1. ASSESSING EDUCATIONAL ACHIEVEMENT

When the United States Office of Education was founded in 1867, one charge set before its commissioner was to determine the nation's progress in education. Almost 100 years later, the National Assessment of Educational Progress began charting educational change under the guidance of Dr. Ralph Tyler and Dr. Francis Keppel. The project is now beginning its sixth year of field assessment.

National Assessment is a project of the Education Commission of the States and was established to measure educational outcomes.

---

*National Assessments of Science, 1969 and 1973: A Capsule Description of Changes in Science Achievement, Science Report No. 04-S-00, February 1975. This report will be released in February, 1975.

The project's goal is to provide reliable information describing what young Americans (at the ages of 9, 13, 17 and 26 to 35) know and can do. Specifically, the assessment is designed (1) to obtain census-like data on knowledge, skills, and attitudes at regular intervals and (2) to measure the growth or decline in educational attainments.

National Assessment has completed a first assessment in Citizenship, Reading, Literature, Music, Social Studies, Mathematics and Career and Occupational Development, as well as second assessments in Science and Writing. The tasks included in each assessment have been judged important by representative panels of scholars, laypersons and educators and these tasks represent things that should be taught in American schools.

The four age levels NAEP measures essentially mark the end of primary, intermediate, secondary and post-secondary education. For each age NAEP reports region, sex, color, level of parents' education and size-and-type of community. These reporting groups were selected because they reflect areas where differences in achievement occur in the population: Region and sex groups have traditionally shown large differences in educational attainments; schools are thought to vary with the size and type of community (STOC) they serve; and color and level of parental education (PED) are believed to differentiate socio-economic and home and family environments.

NAEP uses an estimate of the percent of people who can perform a task or group of tasks as a measure of educational attainment.

NAEP reports relative group performances as "group effects." The group effect is the difference between the percentage of a group and the percentage of all people in the nation at that age who can perform a task or answer a question correctly.

National Assessment does not conduct longitudinal studies in the classical sense but rather monitors performance for a particular age group at successive points in time. This approach has both advantages and limitations. Its chief advantage is that it does allow a comparison of how well thirteen-year-olds of today, for example, can perform the same tasks that thirteen-year-olds were asked to perform three, four, or five years earlier.

Our first overall change measure in Science showed a decline in achievement. The purpose of this paper will be to explicate a general problem associated with change measures of the type National Assessment uses and to explore alternate approaches which might better illuminate possible factors associated with observed results.

For any single point in time, observed group effects estimate relative levels of achievement for subpopulations such as Southeast or Blacks as they exist in our country. Interpretations of observed group percentages, however, can be misleading in several ways. The fact that a group's relative performance is labeled as a Northeast or Southeast regional "effect" does not mean that differences in these effects occur solely because the respondents live in the Northeast or Southeast. For example, a large fraction of respondents in the Northeast live in large cities while a larger fraction of respondents in the Southeast live in rural areas. Consequently,

size-and-type of community effects may be masquerading as part of an observed regional effect. Similarly, persons whose parents went beyond high school are more frequent in affluent communities than in the country as a whole and persons whose parents had no high school are more frequent in extreme rural communities. In this case parental education effects may be masquerading as size-and-type of community effects.

Confusion about group effects due to masquerading arises when the mixture of characteristics is unbalanced from one group to another. Since most of the groups are distributed disproportionately in the national population, our weighted probability sample automatically preserves this imbalance in our percentage estimates.

Thus, we have a problem in understanding what the data have to tell us about the relative achievements among subpopulations at one point in time. A comparison of achievement levels at two points in time becomes more complex since a change measure is affected not only by a group's changing performance but also by change in the inherent subpopulation distributions.

Demands are continually placed on education to find solutions to difficult problems and provide a better education for everyone. At the same time the pace of changing views and emerging opinion about the nature and solutions to problems facing education in our nation far outdistances the aggregation of supporting data. Unfortunately, this climate increases the tendency to go beyond the capability of observational data and decreases resistance to making inferences about causes which are at best uncertainties.

Much has been written about the differences between observation and experimentation and the danger of inferring cause from observational studies. Not much has been said, however, about the need to combine skilled observation with incisive analysis in order to generate suggestions for experimental studies. In his lecture on social experimentation, Frederick Mosteller [3] pointed out that "...we need both mechanisms, one to generate suggestions that might lead to improvement, the other to eliminate most suggestions as ineffective."

Analysis and reanalysis of observational data cannot dispel uncertainties about causes, but they can help us gain new perspectives, generate new hypotheses and perhaps attain a better understanding of the data. For example, a statistical adjustment which balances the distribution of group characteristics may lead to a better understanding of differences among group effects. This is the basic purpose for performing a "balanced analysis" of National Assessment data.

As we shall see, this is much easier to do for single measures of Science achievement than for an analysis of change in Science achievement over time.

2. THE BALANCING PROCEDURE FOR ONE POINT IN TIME

Sample survey data like the kind National Assessment has been collecting over the last five years is frequently adjusted in the sense of forcing sample frequencies or sample ratios to agree with population figures that are known from other sources. This is a desirable procedure because presumably the adjusted sample

represents the populations better and sampling variability of the adjusted frequencies is reduced to some extent.

As a result of this adjustment, the disproportionate distributions are not greatly altered, merely refined. For our purpose, however, we need to go beyond this. It is easier to think about a single marginal group effect if the distributions of the other factors in that group are represented in the same proportions as in the whole age population. Then a direct comparison of any two marginal groups is unconfounded with differing mixtures of these other variables.

Many analyses or data adjustments are possible and specific concerns often direct choices for adjusting data. Tukey [4] specified a data adjustment procedure which defined marginal effects to be non-sensitive to simple disproportionate distributions of other variables. For reasons which favor reader/listener understandability rather than statistical efficiency Tukey chose to fit an additive linear model to the margins. Employing the usual least squares procedures led to the "conditions for balance" where the observed number of successes equals the fitted (balanced) number of successes for each marginal group. These conditions for balance were first written by Tukey as follows:

$$\Sigma n_{ijk\ell m}(P_{nat} + \Delta P_i + \Delta P_j + \Delta P_k + \Delta P_\ell + \Delta P_m) = \Sigma C_{ijk\ell m}$$

where the sum above is taken in turn over all indices except one thereby generating a set of 21 equations. Each equation corresponds to a group denoted by one value of the indices $i = 1, \ldots, 4$; $j = 1,2$; $k = 1, \ldots, 7$; $\ell = 1,2,3$; and $m = 1, \ldots, 5$ belonging to

Region, sex, STOC, color and PED respectively, and where:

P_{nat} is the overall national percent correct for the age group;

ΔP is the balanced group effect corresponding to each value of the indices;

$n_{ijk\ell m}$ is the weighted number of observations in each 5-way cell; and

$C_{ijk\ell m}$ is the weighted number of correct responses in each 5-way cell.

A solution to the 21 simultaneous equations gives a set of fitted group effects. These effects were designed so that when added together and with the national percentage and when multiplied by the actual number of cases, they would give a fitted number of successes equal to the observed number of successes. This set of equations is, however, not of full rank and cannot yield unique balanced ΔP's directly. The number of linearly independent equations is 16 but can be increased by appropriately replacing 5 of the 21 equations with the following usual side conditions imposed when an additive linear model is fitted to a multi-way crossed classification

$$\sum_i n_{i\cdots}\Delta P_i = \sum_j n_{\cdot j\cdots}\Delta P_j = \sum_k n_{\cdot\cdot k\cdot\cdot}\Delta P_k = \sum_\ell n_{\cdots\ell\cdot}\Delta P_\ell = \sum_m n_{\cdots\cdot m}\Delta P_m = 0$$

of the data where the dot "$\cdot$" notation denotes the sum over the replaced subscript. A solution of the independent set of equations results in a unique set of balanced group ΔP's.

The balancing equations can be shown to be algebraically equivalent to the usual set of normal equations resulting from

minimization of the error sum of squares for a five factor additive linear model

$$Y_{ijk\ell m} = \hat{P}_{nat} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\theta}_\ell + \hat{\phi}_m + e_{ijk\ell mr}$$

where $Y_{ijk\ell m}$ is the weighted response for the r-th person in the ijkℓmr-th cell, and α, β, γ, θ, and ϕ correspond to the ΔP's with the same subscript and eijkℓmr is random error associated with the r-th person. The solution to the normal equations provides a simple least squares estimates of the group effects. If the variability of the percentages in the five-way cell combinations happen to be proportional to the reciprocal of the corresponding cell weights, then these group effects are minimum variance estimates. Since the original conditions for balance are preserved in the algebraic equivalence, balanced group effects are also least square estimates.

The aim of balancing observed group effects at one point in time is to try and remove masquerading due to the existing disproportionate distribution of other sub-populations at that point in time. Over time, the distribution of sub-populations shift and the extent of the masquerading of one group as another also changes. Thus, observed changes in performance over time are confounded by the effects of changes in masquerading due to shifting distributions of sub-populations. The next section explores a simple way of adjusting for the effect of sub-population shifts.

3. ADJUSTMENTS FOR THE IMPACT OF SUB-POPULATION SHIFTS

Over a period of years the composition of the population is gradually changing. More students now have parents in the upper educational groups. Size and type of community proportions are changing as are racial proportions. Also, not only are proportions of these groups shifting over time, but they are shifting differentially within other groups. To better understand the nature of change in achievement levels it is helpful to ask: How much of the observed change in achievement is due to changes in relative proportions of the various groups and changes in disproportionalities within each other?

Balancing at one point in time, as described in Section 2, is a way of estimating a group's performance after adjusting sub-population proportions within that group to marginal population proportions. For change in performance over time, we want to adjust the shift in sub-population proportions. One way is to adjust population proportions in the second cycle to correspond identically to population proportions in the first cycle.

Adjusting Changes in National Performance

$P_{nat_2} - P_{nat_1}$ represents the observed change between cycles one and two in national achievement (as measured by percentage of an age group responding acceptably to a task). In order to study the effects of changing disproportionalities, it is helpful to introduce some additional notation. An expected cell percent correct for cycle one can be expressed in terms of the balanced group effects as follows:

$$P_{ijk\ell m_1} = P_{nat_1} + \Delta P_{i_1} + \Delta P_{j_1} + \Delta P_{k_1} + \Delta P_{\ell_1} + \Delta P_{m_1}$$

then,

$$P_{nat_1} = \sum_{ijk\ell m} n_{ijk\ell m_1} P_{ijk\ell m_1} / n_{.....1}$$

and P_{nat_2} can be expressed similarly as

$$P_{nat_2} = \sum_{ijk\ell m} n_{ijk\ell m_2} P_{ijk\ell m_2} / n_{.....2} .$$

Now let P'_{nat_2} represent what the estimated national performance would be in cycle two if the balanced or expected group effects had not changed

$$P'_{nat_2} = \sum_{ijk\ell m} n_{ijk\ell m_2} P_{ijk\ell m_1} / n_{.....2} .$$

Then, $P'_{nat_2} - P_{nat_1}$ is a simple estimate of the impact of shifting population proportions on the national performance assuming the absence of real changes in sub-group performance over time.

Adjusting Changes in Group Performance

For the first cycle an observed performance for group i can be expressed as:

$$P_{i....1} = \sum_{jk\ell m} n_{ijk\ell m_1} P_{ijk\ell m_1} / n_{i....1}$$

and for cycle two as

$$P_{i....2} = \sum_{jk\ell m} n_{ijk\ell m_2} P_{ijk\ell m_2} / n_{i....2} .$$

Then the observed difference in performance between cycles for group i is

$$P_{i....\,2} - P_{i....\,1} .$$

Now, let $P'_{i.....2}$ represent what the estimated group performance would be in cycle two if the balanced or expected group effects had not changed,

$$P'_{i....\,2} = \Sigma\Sigma\Sigma\Sigma_{jk\ell m}\, n_{ijk\ell m_2}\, P_{ijk\ell m_1} / n_{i....\,2}$$

then $P'_{nat_2} - P_{nat_1}$ estimates the impact of shifting sub-population proportions on group performance assuming the absence of any real change in sub-group performance.

A Simple Illustrative Example

Table 1 demonstrates changing proportions between cycle one and cycle two for 13-year-olds for groups defined by race and region only. The total population of in-school, English-speaking 13-year-olds increased from an estimated 3,380,298 in cycle one to an estimated 3,793,959 in cycle four. These estimates are based on probability samples of approximately 25,000 13-year-olds each year.

Table 2 presents estimates of the two-way cell performances P_{ij} where $P_{ij} = P_{nat} + \Delta P_i + \Delta P_j$, i = 1,4 regions, j = 1,3 race categories. P_{ij}'s are shown for each year on one exercise concerned with the consequences to man of the elimination of all green plants. It should be remembered that in this simplified example the category labels represent much more than the label

would appear to stand for. For example, the label "Black" includes the tendency for more Blacks to have low parental education, low socio-economic status, less schooling than Whites of the same age and more Blacks come from larger families--which a recent study indicated to be correlated negatively with achievement. [1] Undoubtedly other variables related to the lower Black performance represent the net result of having been disadvantaged as a group for a long time.

By applying the previous adjustment formulas for shifting populations, adjusted cycle two performances are obtained for the National and the Northeast and Southeast regions and shown in Table 3. The observed differences show that the Southeast declined (-1.56 percentage points), the National decline was (-8.43 percentage points), and the Northeast declined (-11.64 percentage points). The effect of the shifting population proportions is most evident in the Southeast where the adjusted performance ($P'_{2.2}$) is two percentage points higher than actually observed. This implies that the observed change is 2.00 percentage points better than would be expected because of the favorable shift in relative proportions of the race categories actually occurring in the Southeast. With no change in sub-population proportions we would expect to observe a change for the Southeast of -3.65 rather than the observed -1.56.

4. CONCLUSIONS

We have illustrated two simple kinds of data adjustments directed at science achievement measures. One is aimed at removing

masquerading and double counting due to disproportionate distributions of sub-populations. The other is aimed at estimating the impact of shifts of these distributions over time and the corresponding shifts in masquerading which might be confounded with changes in real performance over time.

Statistically, more sophisticated data adjustments could have been employed. But even these simple adjustments give us some understanding about the complex nature of the differences in science achievement among groups and how the nature of these differences becomes even more complex as achievement changes over time.

National Assessment's major goal is to provide reliable data about the changing quality and inequality of achievement in this country. Our observed performances and changes in performance can document these facts very well. What we have done so far is really just the beginning. In order to go beyond the observed data and obtain some notions about the mechanisms involving the complex set of factors affecting education, we need to do more. We need to measure other factors as well as measure better the ones we already have. Finally, we need better and more well-conceived data adjustments for the full value of adjusting data depends on careful, clear documentation of the limited steps we have taken, what we have learned, and what we still don't know.

REFERENCES

[1] Chance, Paul, "Birth Order and Intelligence," Psychology Today, January 1975, Vol. 8, No. 8.

Table 1. Proportions of Race and Regional Subgroups as a Percent of All 13-year-olds

Cycle 1 (1969)

	Race			
	White	Black	Other	All
Region				
Northeast	22.57	2.12	1.04	25.73
Southeast	15.51	6.62	0.53	22.66
Central	26.69	2.89	0.72	30.30
West	15.98	1.61	3.72	21.31
All	80.75	13.24	6.01	100.00

Cycle 2 (1973)

	Race			
	White	Black	Other	All
Region				
Northeast	20.99	2.64	0.83	24.46
Southeast	16.95	5.21	0.26	22.42
Central	24.65	3.01	0.66	28.32
West	18.11	1.78	4.91	24.80
All	80.70	12.64	6.66	100.00

Table 2. Estimated Performance (P_{ij}) for Two-way Cells on One Exercise*

Cycle 1 (1969)

Region	Race White	Race Black	Race Other	All
Northeast	53.50	23.86	33.71	50.26
Southeast	51.91	22.27	32.12	42.79
Central	58.96	29.32	39.17	51.11
West	57.26	27.62	37.47	51.56
All	55.74	24.71	36.55	50.48

Cycle 2 (1973)

Region	Race White	Race Black	Race Other	All
Northeast	42.71	12.58	18.01	38.62
Southeast	48.52	18.39	23.82	41.23
Central	48.26	18.13	23.56	44.48
West	50.45	20.32	25.75	43.40
All	47.36	17.39	24.49	42.05

*Exercise # 201071, 13-year-olds.

Table 3. Observed and Adjusted Percentages for Regional Groups for One Exercise*

	$P_{i\cdot 1}$	$P_{i\cdot 2}$	$P_{i\cdot 2} - P_{i\cdot 1}$	$P'_{i\cdot 2}$	$P'_{i\cdot 2} - P_{i\cdot 1}$	(3) - (5)
	(1)	(2)	(3)	(4)	(5)	(6)
National	50.48	42.05	- 8.43	50.56	+ 0.08	- 8.51
Northeast	50.26	38.62	-11.64	49.63	- 0.63	-11.01
Southeast	42.79	41.23	- 1.56	44.79	+ 2.00	- 3.56
Central	55.66	44.48	-11.18	55.35	- 0.31	-10.87
West	51.56	43.40	- 8.16	51.21	- 0.35	- 7.81

*Exercise # 201071, 13-year-olds